

Age estimation from a face image in a selected gender-race group based on ranked local binary patterns

Andrey Rybintsev¹

Received: 28 July 2016 / Accepted: 6 February 2017 / Published online: 21 February 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract An algorithm to classify people by age from face images based on a two-stage support vector regression is proposed. Only the most significant local binary patterns are used as descriptive features of an image. The distinctive feature of the proposed approach is in the use of a sequential procedure that involves classifying images of people first by gender, then by race in each gender group and only then by age within a selected gender-race group. In order to increase the accuracy of the classification, a bootstrapping procedure (learning on “hard” examples) is used at each stage. The use of this approach has made it possible to improve an accuracy of the classification by 12% for gender, by 15% for race and by 2 years for age (the Mean Absolute Error metric) in comparison with other known algorithms.

Keywords Machine learning · Image classification · Age classification · Local binary patterns · Adaboost · Support vector machine · Bootstrapping · Support vector regression

Introduction

The interest in automatic age classification of people from face image has been growing steadily in recent years. One of the main reasons for this is that the number of human images published on the Internet keeps increasing, and such images need to be annotated automatically for the purposes of filtering results generated by search engines. Automatic

classification by age is also required whenever it is necessary to identify the age of an audience, e.g. to analyze the effectiveness of advertising. Such a classification may also be useful for the purposes of creating human–computer interfaces where the system behavior is adjusted to a specific user based on a number of factors including his or her age.

This article proposed a new algorithm that develops existing approaches to estimating human age [8] with the use of a two-stage scheme of support vector regression. The suggested modifications are based on generating a feature description of an image through applying local binary patterns, identifying only the most relevant (boosted) features, gradually classifying images of people first by gender, then by race in each gender group and only then by age within a selected gender-race group. It is also proposed to apply a bootstrapping procedure that involves learning on “hard” examples with floating rather than fixed age range limits at the second stage of regression.

Related works

Traditionally, an algorithm of age classification of face images involves three stages, i.e. first the face image is normalized, then the feature vector is calculated and finally, classification is performed.

At the normalization stage, the face image is rotated, scaled and cropped in such a way as to ensure proper eye location, i.e. to make sure that the eyes (corners or centers of the eyes) are aligned horizontally. Pixel intensities may be used directly as image features, usually with subsequent dimensionality reduction, e.g. by applying such methods as *Locality Preserving Projections* [9]; geometric features such as the distance between anthropometric landmarks [24]; *Active Appearance Model* parameters [6], and *Local Binary*

✉ Andrey Rybintsev
Andrew.Rybintsev@gmail.com

¹ National Research University “Moscow Power Engineering Institute”, Moscow, Russia

Patterns [19,28,30]. The article [11] proposes so-called *Biologically Inspired Features*. For classification, either a regression method that estimates a person's actual age or a multiple classification method [12] that predicts a person's age group is normally used. Neural networks [16, 17] and random forests [24] are employed as classification algorithms. The support vector machine (SVM) [2,3,11,12] and support vector regression (SVR) [8,30] have gained the widest acceptance.

In [2] the task of age classification is reduced to a number of binary classifications. For each age, a classifier is created that identifies whether the person in the test image is older than the specified age threshold. The person's age is determined as the number of classifiers with positive values. It used support vector machines, and the binary classifiers differed only by a shift of the separating hyperplane, whereas in [3] the binary classifiers learned independently of one another and each classifier could have its own kernel. In [26] this algorithm was applied together with local binary patterns to describe images through features.

Another highly productive idea was to use cumulative features [6] that take into account the relationship between adjacent age groups, thus leading to a higher precision of age identification unlike [3].

Authors of [8] apply a two-stage scheme to identify age based on regression analysis. At the first stage, a regression is constructed for the entire training set to crudely identify the age group, and at the second stage a regression is constructed for the training set within each pre-defined age group.

The authors of article [12] propose first to classify people by gender and approximate age range and then apply a specialized age classifier for the selected gender group. The authors of the article demonstrated that such a sequential procedure notably increases the precision of the final age classification.

The idea of sequential classification appears to be quite promising. This article proposes to expand this approach and apply it not only to preliminary gender but to the preliminary race classification as well. Gender classification is a typical binary classification. Its main methods include support vector machine [14,22,23] and boosting [1,18]. Image pixels [12, 23], Haar-like features [18] and local binary patterns [14,28] are used as features. The studies [10,15,28] are of the utmost interest, as the precision of the algorithms described in them was experimentally assessed based on face images from the Internet.

Race classification is a multiple classification that, as a rule, identifies Caucasians, Asians and Blacks. As a multiple classification is traditionally reduced to a sequence of binary ones, the approaches mentioned above are also applicable to it.

Thus, the following approaches are the most promising for further development:

- using local binary patterns to form an image feature vector;
- sequential gender, race and age classification;
- using a two-stage scheme for age classification that is based on support vector regression using cumulative features.

This article does not consider the algorithms of recognizing people's faces on the general image and their subsequent normalization (size unification, alignment by eye level, etc.), i.e. it is assumed that the algorithm receives already normalized images as input. The above-mentioned procedures are doubtlessly very important but do not constitute the subject of the present research.

Proposed modifications

Modifying the method of the image feature vector formation

Local binary patterns (LBPs) are proposed for use as image features [21]. A LBP is a description of an image pixel neighborhood represented as a binary code. The basic LBP operator applied to an image pixel considers eight surrounding pixels with the intensity of the central pixel used as the threshold. Pixels with an intensity value equal or higher than that of the central pixel are assigned the value of "1", and the rest of the pixels are "0". Thus applying the basic LBP operator to an image pixel results in an eight-bit binary code that describes the neighborhood of the pixel. Subsequently, this code is treated as binary notation of a certain number which is assigned to this pixel. The histogram of LBP values forms the image feature vector.

Face images may be treated as a set of various local features that are effectively described by LBPs; however, a histogram of the entire image encodes only the presence or absence of certain local features but does not contain any information about their location on the image. To account for such information, the image is split into subareas and a special LBP histogram is calculated for each of such subareas. Once concatenated, such histograms form a common histogram that regards both local and global features of the image.

What makes it difficult to use this feature-based image description is that the dimension of feature space is very large. For instance, if a face image is split into $6 \times 7 = 42$ areas, the number of dimensions will be $256 \times 42 = 10,752$. The authors of [25] demonstrated that not all binary patterns have the same informational value. The local image features that are relevant for the purposes of classification are part of *uniform* LBPs that include patterns formed by not more than three series of "0s" and "1s" (e.g. 00000000, 00111000 and

11100001), as such series encode the ends of lines, corners, spots and other specific features of the image. As an eight-bit code ($P = 8$) in total contains $P(P - 1) + 2 = 58$ such binary combinations instead of 256, the total dimension of the feature vector decreases to $58 \times 42 = 2436$, which is much less than the initial; however, this number is still large.

Let us consider other options for reducing the number of dimensions by using a priori information about human face images. To successfully address this task, let us use the following basic assumptions:

1. face symmetry;
2. different information value of various face image areas.

The first assumption is quite obvious; however, the literature does not describe cases where this a priori information was used. When the algorithm of forming the feature space using the LBP method is analyzed, it becomes obvious that the final histogram contains virtually identical pairs of histograms generated for the symmetrical face areas. Figure 1 shows an example of a face image divided into areas that demonstrates the feature-based description similarity of the left and right sides of the face.

This is why deleting half of the image (half of the histograms from the feature description) is not going to affect the informational value of the feature vector (it is unlikely that in a face image values of gender, race or age of the left side would be different to those of the right side). The use of this evident fact causes a halving of the dimensionality of the feature space, i.e. if only uniform binary patterns are considered and if the source image is split into 42 regions, the number of dimensions decreases from 2436 to 1218 as the number of regions is 21.



Fig. 1 An image with examples of virtually identical LBP regions (20 and 23, 33 and 35)



Fig. 2 Generation of a frontal view of the face from the real one [15]

There is another advantage in using a priori information on face symmetry. At the stage of normalization a face image undergoes affine transformation to produce a frontal image from a real one (which is almost always rotated to a certain extent). The biggest problems with image quality that arise in the course of this procedure are related to the reconstruction of the far (partially hidden) half of the face (Fig. 2). The proposed idea of using only half of the image renders redundant the procedure of reconstructing the second half of the face, which saves time on image normalization.

Another condition for reducing the dimensions of the feature space (different information value of face image areas) also seems to be quite obvious (e.g. the image area that depicts an eye or a mouth has much higher information value than the area containing a cheek). Even inside a region there are areas that contain different amount of information hidden in the binary patterns (this fact is partly reflected in the uniform binary patterns). This is why it is recommended to delete from the final feature vector the features that do not affect the final result of the classification. To identify the meaningful features it is proposed to use the *AdaBoost* [27] algorithm that enables training a committee of simple classifiers.

Let us consider a training set for gender classification G^l defined by a multitude of precedent pairs $\{\bar{x}_i, y_i\}$, $i = 1, n$; $G^l \in \mathbb{R}^m$, \bar{x}_i is the image feature vector i of the training set G^l obtained using half of the source image, and y_i is the corresponding value of gender, $y_i \in \{-1, +1\}$. Let us enumerate all features from 1 to 1218, i.e.

$$\bar{x}_i = (x_i^1, x_i^2, x_i^3 \dots x_i^{1218}).$$

Now x_i^j is the frequency of occurrence of the j th sequential feature ($j = 1.1218$) of a uniform binary pattern in the full vector representation of half of the i th image.

In order to identify the most informative of original features it is proposed to use the *Adaboost* [27] algorithm for binary classification. The main idea is to apply a boosting procedure to each of the features, i.e. to force the *Adaboost* algorithm to make decisions on assigning the image to one

class or another separately by each singular feature x_i^j , $i = 1..n$; $j = 1..1218$. In this case the task of classification by means of the **Adaboost** algorithm may be formulated in the following manner:

$$f(x^j) = \text{sign} \left(\sum_{i=1}^n \alpha_i^j b(x_i^j) \right), \quad (1)$$

i.e. the decision function will be generated in such a manner as to classify images only by the j th feature ($j = 1..1218$) of the whole training set G^l . In this relationship the weight coefficients α^j determine the significance of the basic function $b(x)$. It is proposed to use the simplest binary classifiers with separation by threshold value (*Decision Stump*) to represent this function. This threshold value is calculated as the average value of the x_i^j parameter of the whole training set G^l .

Let us define the loss function $L(f(x^j), G^l)$ as the number of errors committed by the committee of 1218 basic decision functions (by the number of features) on features x^j of all objects of the training set G^l :

$$L(f(x^j), G^l) = \sum_{j=1}^n \frac{1 - z_j}{2};$$

$$z_j = y_j \sum_{i=1}^n \alpha_i^j b_i(x_i^j). \quad (2)$$

In accordance with the **Adaboost** algorithm, instead of the threshold function $z < 0$ we will use its continuously differentiable upper bound estimate in the form of $E(z) = \exp(-z)$:

$$L(f(x^j), G^l) \leq \tilde{L}(f(x^j), G^l) = \sum_{i=1}^n \exp(-\alpha_i^j b(x_i^j) y_i).$$

Minimization of this function leads to the following result for the weight coefficients [27]:

$$\alpha^j = \frac{1}{2} \ln \frac{P(x^j)}{N(x^j)}, \quad (3)$$

where $P(x^j)$ is the fraction of correct classifications and $N(x^j)$ is the fraction of erroneous classifications x^j of all objects of the training set: $P(x^j) > N(x^j)$, $P(x^j) + N(x^j) = 1$. For features x^j characterized by $P(x^j) < N(x^j)$, the coefficients $\alpha^j = 0$. The calculated weight coefficients are normalized in such a way that their sum equals 1.

As a result of application of this approach we obtain a combination of weight coefficients α^j that determine the significance of the basic algorithm $b(x)$. Because the basic

algorithm is applied to a single feature x^j , this weight coefficient will characterize the significance of this particular feature x^j when the decision is made on assigning this object to a particular class.

Next, it is proposed to arrange the features by decreasing weight coefficients, introduce the significance threshold δ and employ for the purposes of classification only those features for which the following condition is fulfilled:

$$\sum_{j=1}^N \alpha^j < \delta, \quad (4)$$

i.e. leave only the first N features with the summary weight of δ . These features will be called *boosted* and labeled as $\hat{\mathbf{X}}$, and the formalized training set containing only boosted features will be labeled as \hat{G}^l .

The use of a priori information about the human face along with determining the significance of the image features based on the boosting procedure allows reducing the number of dimensions of the LBP feature space to approximately 250 (with the significance threshold $\delta = 0.95$). These features are proposed for further use in gender, race and age classification.

Modifying the procedure for classifying images of human faces by gender and race

The classification of images by the “race” and “gender” attributes is binary or may be reduced to a binary one. This section proposes a modified support vector machine (SVM) that has achieved the best results to-day by the *Accuracy* [20,22] metric when determining the “gender” and “race” attributes. A soft margin SVM [29] was used as the initial algorithm for modification. This method allows the algorithm to commit errors (the case of linearly indivisible sets) but the magnitude of the errors is minimized.

For the sake of improving binary classification, it is proposed to apply a bootstrapping procedure that involves learning on so-called “hard” examples [7,19]. It should be noted that this procedure is commonly used to solve another problem of computer vision, namely detecting a specified object on an image that contains a number of various (similar) objects [19] (e.g. detecting a pedestrian on an image of a street with many people in cars, on bicycles and on billboards). The main problem in solving this task is a huge number of background examples (non-pedestrians) as compared to the target objects. This is why the central idea of bootstrapping is that the training set should contain “hard” examples, i.e. examples of background that may be erroneously identified by the classifier as the target objects.

Let us illustrate the application of the bootstrapping procedure with an example of gender classification. Let us assume

we have a training set \hat{G}^l defined by a set of precedent pairs $\{\hat{x}_i, y_i\}$, $i = 1, n$; $\hat{x}_i \in \mathbb{R}^m$, $y_i \in \{-1, +1\}$, where the image feature vector i contains only boosted features that have been generated using the procedure described above.

The proposed bootstrapping procedure consists of three steps:

1. The training set \hat{G}^l is randomly divided into two parts, \hat{G}_1^l and \hat{G}_2^l , in the ratio of 1 to 2: $\hat{G}_1^l \cap \hat{G}_2^l = \emptyset$. The first part \hat{G}_1^l is used as a training set for generation of a preliminary estimation \tilde{f} of the decision function f using a soft-margin SVM.
2. The second part \hat{G}_2^l is used as a test set to evaluate the predictive ability of the decision function \tilde{f} . In the course of this procedure all wrongly classified objects $\hat{x}_i \in \hat{G}_2^l$ are identified.
3. Erroneously classified objects identified in step 2 (examples that are hard to classify) are added to the training set \hat{G}_1^l , i.e. a new training set \hat{G}_{new}^l enhanced with “hard” examples is generated. This enhanced set is used for retraining, i.e. the final version of the decision function f is generated.

The proposed modifications (using only a half of the image, selecting the most relevant LBP features using *Adaboost* and bootstrapping at the stage of classifier training based on a soft margin SVM) result in forming of a classifier that is able to classify human face images by gender.

To identify the “race” attribute in a face image, a multiple classification is used under the “one versus all” scheme, where the objects of the “Caucasian” class are identified first based on the proposed modification of the binary classification, and then the remaining objects are classified as either “Asian” or “Black”.

As a result of the proposed method the following set of classifiers is generated:

- a gender classifier (assigns labels “M” and “F” to images);
- a classifier by race among men that identifies Caucasians among the images that carry the “M” label, i.e. the images receive the label “CM”;
- a classifier by race among men that identifies Asians among the images that are labeled “M” and have not been classified as Caucasian, i.e. the images receive the label “MM”, and the rest of the images are classified as Black and receive the label BM;
- a classifier by race among women that identifies Caucasians among the images that carry the “F” label, i.e. the images receive the label “CF”;
- a classifier by race among women that identifies Asians among the images that are labeled “F” and have not been

classified as Caucasian, i.e. the images receive the label “MF”, and the rest of the images are classified as Black and receive the label BF.

The result of these classifiers is the attribution of one of the following labels to the images: CM, MM, BM, CF, MF, BF.

Further the method of age classification may be applied to each separate gender-race group of the images.

Modification of the two-stage scheme for human age classification

Based on an analysis of various approaches to age classification by face images using regression analysis, the following action sequence that combines the best of currently available ideas is proposed:

- using support vector regression (SVR) [29];
- using the idea of *cumulative* features [4] as the source data for the SVR;
- using a two-stage scheme of determining the “age” attribute [8].

The basic idea of cumulative features is as follows:

Let us consider the X^l training set of size n that is defined by a multitude of precedent pairs $\{\bar{x}_i, y_i\}$, where \bar{x}_i is the vector of “boosted” features of image i , and y_i is the corresponding age value. For each precedent, the scalar value of age is transformed into a vector of cumulative features \bar{a}_i with m dimensions that corresponds to the range of ages (y_{\min}, y_{\max}):

$$m = y_{\max} - y_{\min} + 1.$$

Each element of the cumulative feature vector is determined as follows:

$$a_i^j = \begin{cases} 1, & j \leq y_i - y_{\min}; \\ 0, & j > y_i - y_{\min}. \end{cases}$$

In this case the cumulative feature vector will be assigned ($y_i - y_{\min}$) the value of 1, whereas all the others will be deemed equal to 0.

The task of finding a correlation between the original boosted and cumulative features is reduced to the task of ridge regression and is solved for each gender group separately; however, unlike the basic algorithm [4], all the image features are not used to identify the cumulative features, but only those that are most relevant for the classification (the boosted features). This modification provides for substantial

(proportional to the decrease of the dimensionality of the feature space) saving of time required for classification.

The algorithm of determining cumulative features \tilde{a}_i based on the original features is described in detail in [4] and is reduced to a quadratic programming problem.

The basic idea of the two-stage age estimation scheme is as follows. At the first stage, the approximate age is calculated using a decision function $f_0(\hat{x})$ defined on the basis of a regression constructed based on the training set \hat{G}^l across the whole range of ages (y_{\min}, y_{\max}). At the second stage, this value is refined by means of a decision function $f_d(\hat{x})$ generated on the basis of a regression constructed for a specific age group d that includes the value of age estimated at the first stage. As in the publication [8], we shall call the regression used at the first stage *global* and the one used at the second stage *local*.

In accordance with the basic algorithm, the whole age interval (y_{\min}, y_{\max}) is divided into non-intersecting ranges that are used to determine a decision function for each of them on the basis of the local regression. This article proposes rejecting fixed range limits and using floating limits that depend on the specific age value obtained at the first stage. The following statement is the basic premise of the proposed modification: with age identification, the significance of the error depends on the absolute age value. For example, an error of 2 years in estimating the age of a 15-year-old adolescent appears quite significant while an error of 5 years in estimating the age of a 70-year-old man could hardly be considered as such.

At the second stage, it is proposed to employ the range of ages ($y - d^-(y), y + d^+(y)$) for local regression, where y is the approximate value of age estimated at the first stage. It should be noted that in the general case $d^-(y) \neq d^+(y)$, i.e. the range may be asymmetric. Selecting such a range is appropriate only in cases when the distribution of the number of precedents by age is substantially irregular.

The second way of modifying the two-stage system of estimating age is the modification of the loss function used in regression analysis. The methods known today [8] use an ε -sensitive loss function with a fixed ε sensitivity value. In line with the above-mentioned idea of floating range limits, at the second stage of the two-stage age estimation scheme it is proposed to apply different loss function sensitivity values depending on the absolute age, i.e. $\varepsilon = \varepsilon(y)$ depends on the y age value obtained at the first stage. Figure 3 shows a graphic illustration of the idea of the proposed modifications.

To summarize, we propose the following procedure of generating a set of decision functions for estimating age using SVR based on the suggested modifications of the two-stage age classification scheme:

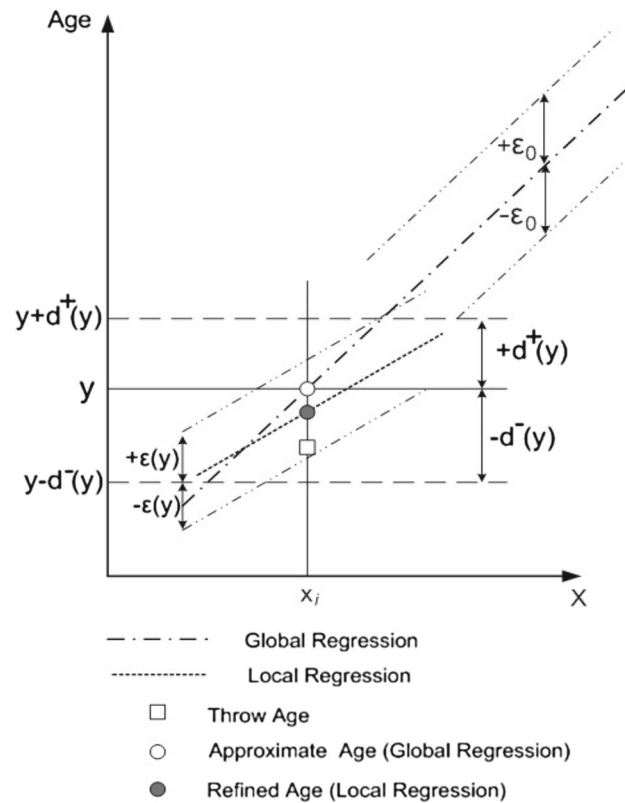
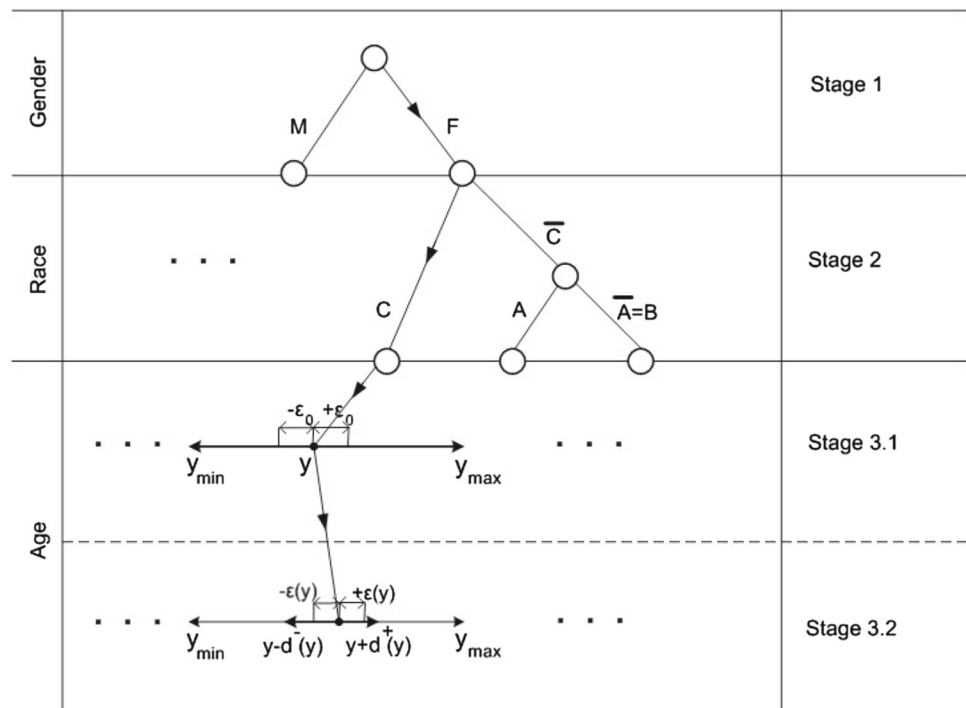


Fig. 3 Illustration of the proposed modifications of the global and local regressions

1. A set of boosted features \hat{X}^l is generated for the formalized training set \hat{G}^l in accordance with the algorithm proposed above.
2. Cumulative features \hat{A}^l are determined in accordance with the method of generating cumulative features based on boosted features \hat{X}^l .
3. The obtained cumulative features \hat{A}^l of the formalized training set \hat{G}^l are employed for regression analysis using SVR for the whole training set with a predefined value of ε -sensitivity ε_0 , i.e. a decision function $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$ is generated for the preliminary estimation of the age value \tilde{y}_i (global regression), i.e. $\tilde{y}_i = f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$, $i = 1, n$. A bootstrapping procedure is used on the whole training set:

- the set \hat{G}^l is randomly divided into two parts \hat{G}_1^l and \hat{G}_2^l with a ratio of 1 to 2;
- a decision function is generated for the set \hat{G}_1^l using SVR;
- the set \hat{G}_2^l is used for training tests with identification of “hard” examples (examples are considered “hard” if they do not fall within the sensitivity band ε_0);

Fig. 4 The general scheme of the algorithm for estimating the age from a face image in a specified gender-race group



- the set \hat{G}_1^l is expanded with “hard” examples, i.e. the set \hat{G}_{1new}^l is generated;
- the set \hat{G}_{1new}^l is used for final determination of the decision function $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$.

- For each value of age y a range of ages $(y - d^-(y), y + d^+(y))$ is selected with offset values that depend on y and are defined by the expert based on the specific features of the problem being solved. The most obvious idea is that the higher the age value, the wider the range; but other rules can be applied as well.
- Inside each range for the corresponding subset of the training set \hat{G}^l : $y_i \in (y - d^-(y), y + d^+(y))$, SVR is employed for regression analysis where the sensitivity of the loss function $\varepsilon(y)$ depends on the value of y and is also defined by the expert based on the specific features of the problem being solved. The most obvious idea is still that the higher the value of y , the lower the sensitivity, i.e. the higher the $\varepsilon(y)$. Thus a decision function $f_y(\hat{a}, \varepsilon(y), d(y), \hat{G}^l)$ is generated for estimating the refined age value. Here, as in step 4, a bootstrapping procedure is used for each range $(y - d^-(y), y + d^+(y))$ of the training set (the examples are still considered “hard” if they fall outside of the sensitivity band $\varepsilon(y)$).

The final result of the proposed algorithm is a decision function $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$ that estimates the approximate age y across the whole range of ages (y_{\min}, y_{\max}) (global regression), and a set of functions $f_y(\hat{a}, \varepsilon(y), d(y), \hat{G}^l)$ that

estimates the refined value of age at the second stage (local regression).

Summary of the proposed algorithm

The combination of the proposed modifications forms a general approach to estimating a person’s age from a face image based on preliminary gender and race classification (Fig. 4). The preliminary stage Generation of “boosted” features \hat{X}^l for the training set of images:

- selection of half of a normalized face image and application of **LBP** to this half-image in order to generate a feature vector containing only *uniform* patterns;
- usage of the **Adaboost** method for the “gender” attribute to exclude insignificant features from the image feature vector (the sum of weight coefficients of the boosted features is equal to 0.95).

Stage 1 Generation of a decision function $f^{\text{gender}}(\hat{x})$ for the binary gender classification the objects using SVM on the boosted features and the proposed bootstrapping procedure.

Stage 2 Generation of decision functions $f_M^{\text{race}}(\hat{x})$ also $f_F^{\text{race}}(\hat{x})$ within each gender group for a multiple classification of objects by race using a binary classification (the “one versus all” scheme) taking into account the proposed bootstrapping procedure. In total, two decision functions are generated for each gender group:

- function $f_F^C(\hat{x})$ identifies the objects belonging to the “Caucasian” class among the objects classified as “Female”;
- function $f_F^{AB}(\hat{x})$ identifies the objects belonging to the “Asian” class among the objects classified as “Female”; the rest of the objects are classified as “Black”;
- function $f_M^C(\hat{x})$ identifies the objects belonging to the “Caucasian” class among the objects classified as “Male”;
- function $f_M^{AB}(\hat{x})$ identifies the objects belonging to the “Asian” class among the objects classified as “Male”; the rest of the objects are classified as “Black”.

Stage 3 Generation of a set of decision functions for the two-stage scheme of age estimation using SVR for each gender-race group:

- the cumulative features are generated \hat{A}^{lp} based on boosted features \hat{X}^{lp} of the training set for each combination of gender (“Male”, “Female”) and race (“Caucasian”, “Asian”, “Black”) groups p ; $p = (1, 6)$;
- the cumulative features \hat{A}^{lp} are used for regression analysis by SVR on the whole training set of the gender-race group p with a predefined value of ε -sensitivity ε_0^p , i.e. a decision function is generated $f_0^p(\hat{a}^p, \varepsilon_0^p)$ (in general, the sensitivity ε of the loss function may depend on the gender-race group p) for a preliminary estimation of age (*stage 3.1, global regression*) for each gender-race group p , $p = (1, 6)$. As a result, six decision functions are generated, each capable of calculating an approximate value of the “age” attribute for each object with the attributes of “gender” and “race” determined at the previous stages using the boosted features of this object. A bootstrapping procedure is used on the entire training set for generating decision functions $f_0^p(\hat{a}^p, \varepsilon_0^p)$ for each gender-race group p ;
- for each age value y^p in each gender-race group p an age range $(y^p - d^{p-}(y^p), y^p + d^{p+}(y^p))$ is selected with values of $d^{p-}(y^p)$ and $d^{p+}(y^p)$ (in general, the breadth of the range may depend on the gender-race group p), that are defined by the expert on the basis of the specific features of the problem being solved; within each range belonging to value y^p , for the corresponding subset of precedents (x_i^p, y_i^p) : $y_i^p \in (y^p - d^{p-}(y^p), y^p + d^{p+}(y^p))$; $p = (1, 6)$; $i = (1, n^p)$. n^p is the number of precedents of the training set in the gender-race group p , SVR is used for regression analysis (*stage 3.2, local regression*) where the sensitivity of the loss function $\varepsilon^p(y^p)$ depends on the value of y^p and is also defined by the expert based on the features of the problem being solved (in general, the sensitivity ε of the loss function may depend on the gender-race group p). As a result, a set of

Table 1 Contents of the training sets

Attribute	Contents
Gender	Images of women and men of different races and ages
Race	Images of men of different races and ages Images of women of different races and ages
Approximate value of age (global regression)	Images of Caucasian men of different ages Images of Caucasian women of different ages Images of Asian men of different ages Images of Asian women of different ages Images of Black men of different ages Images of Black women of different ages
Refined value of the age attribute (local regression)	Images of Caucasian men in the age range $(y_1 - d_1^-, y_1 + d_1^+)$... Images of Caucasian men in the age range $(y_r - d_r^-, y_r + d_r^+)$ Images of Caucasian women in the age range $(y_1 - d_1^-, y_1 + d_1^+)$... Images of Caucasian women in the age range $(y_r - d_r^-, y_r + d_r^+)$ Images of Asian men in the age range $(y_1 - d_1^-, y_1 + d_1^+)$... Images of Asian women in the age range $(y_r - d_r^-, y_r + d_r^+)$

decision functions $f_y^p[\hat{a}^p, \varepsilon^p(y^p), d^{p+}(y^p), d^{p-}(y^p)]$ is generated for the purpose of estimating the refined age values for each gender-race group p and age y^p : $p = (1, 6)$, $y^p \in (y_{\min}, y_{\max})$. A bootstrapping procedure is used for generating a decision function $f_y^p[\hat{a}^p, \varepsilon^p(y^p), d^{p+}(y^p), d^{p-}(y^p)]$ for each gender-race group p in every age range $(y^p - d^{p-}(y^p), y^p + d^{p+}(y^p))$.

Table 1 shows the contents of the training sets for each stage of the algorithm.

Experimental results

Training of classifiers to determine the attributes of a person based on a face image implies the availability of a database of images with labeled values of gender, race and age. Not

Table 2 Available databases

No.	Base of images	Gender	Race	Age
1	FERET	+		
2	MORPH	+		+
3	FG-NET	+		+
4	Mall	+	+	
5	LFW	+	+	+

many of those are publicly available. If a database of faces that do not carry labels for gender or race is available, it is possible to manually process such a database and specify the corresponding attribute values. In case of the “age” attribute it is practically impossible to perform a similar procedure, as a human cannot estimate the age value from a face image.

Table 2 provided below lists the publicly available databases of people’s faces with specification of their parameters.

An analysis of publicly available face image databases has shown that they have limited use for purposes of analysis of accuracy of gender, race and age classification (only the accuracy of some of the attributes may be analyzed). It must be noted that the majority of databases contain images obtained while photographing people specifically, leading to inflated classification quality estimates in comparison with analysis of images obtained in real-life conditions. The most suitable database for purposes of the present analysis is the LFW, but it contains very few images of faces younger than 20 years, and age classification in this particular range is considered the most important in many cases. In view of the above, it was decided to create our own face image database using publicly available sources (social networks) on the Internet.

The task of generation of the face image database is made significantly simpler by the availability of special instruments—services aggregating data from social networks that offer search capability for all available social networks via a single interface (<http://people.yandex.ru>). The simplest age and gender classifiers were employed for preliminary distribution of uploaded images by the corresponding attributes. Then all images obtained were manually viewed to check and refine the gender and race data. This work has been performed by the author in cooperation with the staff of the Computer Graphics Lab of the Moscow State University. As a result, the BigSample database has been generated containing 169,629 face images. Samples of images are shown in Fig. 5.

The race and gender composition of the BigSample image database is presented in Table 3, and the distribution of images by age is shown in Fig. 6.

The results of the research are shown in Tables 4 and 5.

An analysis of the obtained results permits drawing the following conclusions:



Fig. 5 Samples of images from the BigSample database **a** woman, black, 25 years old **b** man, Asian, 71 years old **c** man, Caucasian, 21 years old **d** woman, Caucasian, 52 years old

Table 3 Race and gender composition of the BigSample image database

	Men	Women	Total number of images
Caucasian	50,210	42,419	92,629
Asian	22,740	19,682	42,422
Black	18,045	16,533	34,578
Total number of images	90,995	78,634	169,629

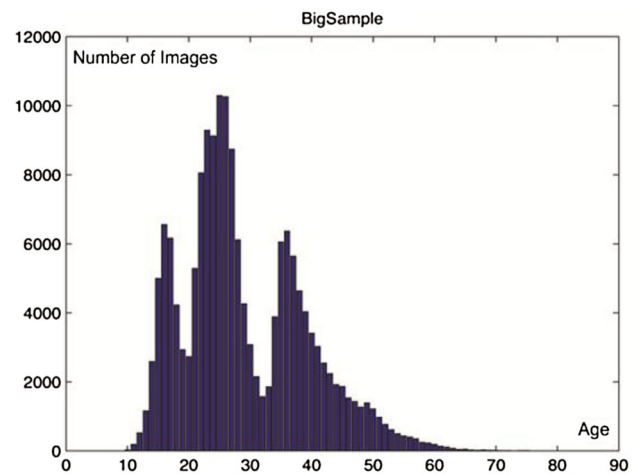


Fig. 6 Distribution of people by age in the BigSample database

- the accuracy of gender classification using a conventional SVM based on LBP on the BigSample image database virtually coincides with the published results [5,22] obtained from actual images from the Internet (the difference in accuracy is lower than 1%); a slightly more accurate gender classification on the FERET and MORTH databases is due to the fact that these databases, unlike BigSample, contain specially prepared images;
- the transition to boosted LBP features does not decrease the accuracy of gender and race classification either in the case of specially prepared databases of face images (FERET, MORPH, Mall), or in the case of real face images from the Internet (BigSample);

Table 4 The results of the research on accuracy of gender classification

Algorithm Features	SVM LBP	SVM Boosted LBP	SVM + bootstrapping Boosted LBP
Accuracy			
FERET (%)	86	86	94
MORPH (%)	87	87	95
BigSample (%)	79	79	92

Table 5 The results of the research on accuracy of race classification

Algorithm Features	SVM LBP	SVM + bootstrapping Boosted LBP	SVM + bootstrapping in a selected gender group Boosted LBP
Accuracy			
Mall (%)	81	90	96
BigSample (%)	74	85	89

Table 6 The results of the analysis of accuracy of age classification

Algorithm Features	Two stages SVR LBP	Two stages SVR + “floating” limits + bootstrapping Boosted LBP	Two stages SVR + “floating” limits + bootstrapping within a predefined gender-race group Boosted LBP	Age classification after gender-race classification Boosted LBP
MAE				
FG-NET	4.65	4.43	4.30	4.31
MORPH	4.61	4.42	4.28	4.29
BigSample	8.07	6.72	6.08	6.09

- the application of a bootstrapping procedure leads to an increase in accuracy of gender and race classification both for databases of specially prepared face images and for databases of real face images from the Internet by $\sim 12\%$ in comparison to the basic method.
- the simultaneous application of boosted LBP features, a bootstrapping procedure and sequential classification of race after the classification of gender lead to an increase in the accuracy of race classification by $\sim 15\%$ in comparison to the basic method.

The analysis of accuracy of age classification employed the most commonly used Mean Absolute Error (MAE) metric, which calculates the average absolute deviation of the age values predicted by the decision function for the objects included in the test set from the true values of age [13, 24, 26]. The results of the analysis of accuracy of age classification are shown in Table 6.

On the basis of results of the analysis, the following conclusions can be made:

- the implementation of the proposed modification of the two-stage regression scheme for age classification in combination with the use of boosted image description

features leads to an increase in accuracy of estimation of this attribute in real face images by about 2 years using the MAE metric, based on Internet face image analysis.

- the higher accuracy of age estimation in face images from FG-NET and MORPH databases is explained by the use of specially prepared face images in these databases
- a slight decrease in the accuracy of age estimation with the use of the proposed method is explained by the fact that sometimes the classifier commits errors in determining the gender or race; this leads to the use of wrong decision function trained on a “foreign” gender-race group at the age estimation stage.

To obtain statistically significant results a k -fold ($k = 5$) cross-validation procedure was applied.

Conclusion

This article proposes an algorithm to estimate human age from face images within a preliminarily selected gender-race group based on local binary patterns used as feature description of an image. To identify the most significant (boosted) features, it is proposed to apply the *Adaboost*

method that resulted in an approximately tenfold decrease in the dimensionality of the feature space. For gender and race classification, the standard approach based on support vector machines is modified by adding a bootstrapping procedure (learning on “hard” examples), while to ensure more precise age estimation it is proposed to combine the idea of cumulative features with two-stage support vector regression with a bootstrapping procedure.

The conducted research has demonstrated that each of the proposed modifications contributes to increased age estimation precision according to the MAE metric. As a result, a 2-year reduction in the average error in age estimation from real face images published on the Internet was achieved as compared to the source algorithm. The obtained results allow recommending this algorithm to solve the task of gender, race and age classification of human face images for use in computer vision systems.

Acknowledgements The study was conducted with financial support from the Russian Foundation for Basic Research (Project No. 17-01-00485a).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baluja S, Rowley H (2007) Boosting sex identification performance. *Comput Vis* 71(1):11–19. doi:[10.1007/s11263-006-8910-9](https://doi.org/10.1007/s11263-006-8910-9)
- Chang KY, Chen CS, Hung YP (2010) A ranking approach for human age estimation based on face images. In: 20th IEEE international conference on pattern recognition, pp 3396–3399. doi:[10.1109/ICPR.2010.829](https://doi.org/10.1109/ICPR.2010.829)
- Chang KY, Chen CS, Hung YP (2011) Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: 2011 IEEE international conference on computer vision and pattern recognition, pp 585–593. doi:[10.1109/CVPR.2011.5995437](https://doi.org/10.1109/CVPR.2011.5995437)
- Chen K, Gong S, Xiang T (2013) Cumulative attribute space for age and crowd density estimation. In: IEEE international conference on computer vision and pattern recognition, pp 2467–2474
- Chen C, Ross A (2011) Evaluation of gender classification methods on thermal and near-infrared face images. In: International joint conference on biometrics (IJCB), pp 367–374. doi:[10.1109/IJCB.2011.6117544](https://doi.org/10.1109/IJCB.2011.6117544)
- Coates T, Edwards G, Taylor C (2001) Active appearance models. *IEEE Trans Pattern Anal Mach Intell* 23(6):681–685. doi:[10.1109/34.927467](https://doi.org/10.1109/34.927467)
- Felzenszwalb P, Gurchik R, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part base models. *IEEE Trans Pattern Anal Mach Intell* 32(9):1627–1645. doi:[10.1109/TPAMI.2009.167](https://doi.org/10.1109/TPAMI.2009.167)
- Guo G, Fu Y, Dyer C, Huang T (2008) Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans Image Process* 17(7):1178–1188. doi:[10.1109/TIP.2008.924280](https://doi.org/10.1109/TIP.2008.924280)
- Fu Y, Xu Y, Huang T (2007) Estimating human ages by manifold analysis of face pictures and regression on aging features. In: IEEE multimedia expo conference, pp 1383–1386. doi:[10.1109/ICME.2007.4284917](https://doi.org/10.1109/ICME.2007.4284917)
- Gao W, Ai H (2009) Face gender classification on consumer images in a multiethnic environment. In: 3d IEEE international conference advances in biometrics, pp 169–178. doi:[10.1007/978-3-642-01793-3_18](https://doi.org/10.1007/978-3-642-01793-3_18)
- Guo G, Mu G, Fu Y, Huang T (2009) Human age estimation using bio-inspired features. In: IEEE international conference on computer vision and pattern recognition, pp 112–119. doi:[10.1109/CVPR.2009.5206681](https://doi.org/10.1109/CVPR.2009.5206681)
- Guo G, Mu G, Fu Y, Dyer C, Huang T (2009) A study on automatic age estimation using a large database. In: IEEE 12th international conference on computer vision, pp 1986–1991. doi:[10.1109/ICCV.2009.5459438](https://doi.org/10.1109/ICCV.2009.5459438)
- Guo G, Zhang C (2014) A study on cross-population age estimation. In: IEEE conference on computer vision and pattern recognition, pp 4257–4263. doi:[10.1109/CVPR.2014.542](https://doi.org/10.1109/CVPR.2014.542)
- Hadid A, Pietikainen M (2009) Combining appearance and motion for face and gender recognition from videos. *Pattern Recognit* 42(11):2818–2827. doi:[10.1016/j.patcog.2009.02.011](https://doi.org/10.1016/j.patcog.2009.02.011)
- Han H, Jain A (2014) Age, gender and race estimation from unconstrained face images. MSU Technical Report: MSU-CSE-14-5
- Konushin VS, Lukina TM, Kuharenko AI, Konushin AS (2013) Person classification upon face image based on simile classifiers. *Syst Means Inform* 23(2):50–61. doi:[10.14357/08696527130204](https://doi.org/10.14357/08696527130204)
- Kuharenko A, Konushin A (2015) Simultaneous classification of several features of a person’s appearance using a deep convolution neural networks. *Pattern Recognit Image Anal* 25(3):461–465. doi:[10.1134/S1054661815030128](https://doi.org/10.1134/S1054661815030128)
- Lapedriza A, Marin-Jimenez M, Vitria J (2006) Gender recognition in non controlled environments. In: ICPR 2006. IEEE 18th international conference on pattern recognition, pp 834–837. doi:[10.1109/ICPR.2006.598](https://doi.org/10.1109/ICPR.2006.598)
- Laptev I (2009) Improvements of object detection using boosted histograms. *Image Vis Comput* 27(5):535–544. doi:[10.5244/C.20.97](https://doi.org/10.5244/C.20.97)
- Ling CX, Huang J, Zhang H (2003) AUC: a statistically consistent and more discriminating measure than accuracy. In: 18th International conference on artificial intelligence, pp 519–524
- Maenpaa T (2003) The local binary pattern approach to texture analysis—extensions and applications. Oulu University Press, Oulu
- Makinen E, Raisamo R (2008) Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Trans Pattern Anal Mach Intell* 30(3):541–548. doi:[10.1109/TPAMI.2007.70800](https://doi.org/10.1109/TPAMI.2007.70800)
- Moghaddam B, Yang M (2002) Learning gender with support faces. *IEEE Trans Pattern Anal Mach Intell* 24(5):707–711. doi:[10.1109/34.1000244](https://doi.org/10.1109/34.1000244)
- Montillo A, Ling H (2009) Age regression from faces using random forests. In: IEEE international conference on image processing, pp 2437–2440
- Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987. doi:[10.1109/TPAMI.2002.1017623](https://doi.org/10.1109/TPAMI.2002.1017623)
- Rybintsev AV, Lukina TM, Konushin VS, Konushin AS (2013) Age estimation upon face image based on local binary patterns and a ranking approach. *Syst Means Inform* 23(2):62–73. doi:[10.14357/08696527130205](https://doi.org/10.14357/08696527130205)
- Schapire R (1999) A brief introduction to boosting. In: 16th International joint conference on artificial intelligence, vol 2, pp 1401–1406

28. Shan C (2012) Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit Lett* 33(4):431–437
29. Vapnik V (1998) *Statistical learning theory*. Wiley, New York
30. Yilionias J, Hadid A, Hong X, Pietikainen M (2013) Age estimation using local binary patterns kernel density estimate. In: *IEEE international conference on image analysis and processing*, pp 141–150. doi:[10.1007/978-3-642-41181-6_15](https://doi.org/10.1007/978-3-642-41181-6_15)